Title: Transposition Cryptography

Brief Overview:

All forms of cryptography belong to one of two families of cryptosystems: substitution or transposition. Thus far, we have studied substitution systems. That is, some element of the plaintext (a letter, a digraph, a syllable, etc) is substituted for some element of cipher. In transposition systems, the plaintext is left unchanged but re-ordered in such a way that if an unintended recipient should get the message and does not know the decryption key, the plaintext would remain unreadable. There is virtually no limit to the number of ways plaintext can be transposed.

However, the easiest and most popular way (probably the most popular because it's the easiest) is matrix transposition. In this form, the message is written into a matrix of predetermined size (# rows, # columns) left-to-right, top-to-bottom, in the normal manner of writing. The plaintext letters are then extracted by columns according to a key. It should be noted that in a matrix of K columns there are K! possible keys. The first column of letters extracted will become the first R letters of cipher, where R equals the number of rows in the matrix.

NCTM 2000 Principles for School Mathematics:

- **Equity:** *Excellence in mathematics education requires equity - high expectations and strong support for all students.*
- **Curriculum:** *A curriculum is more than a collection of activities: it must be coherent, focused on important mathematics, and well articulated across the grades.*
- **Teaching:** *Effective mathematics teaching requires understanding what students know and need to learn and then challenging and supporting them to learn it well.*
- **Learning:** *Students must learn mathematics with understanding, actively building new knowledge from experience and prior knowledge.*
- **Assessment:** *Assessment should support the learning of important mathematics and furnish useful information to both teachers and students.*
- **Technology:** *Technology is essential in teaching and learning mathematics; it influences the mathematics that is taught and enhances students' learning.*

Links to NCTM 2000 Standards:

- **Content Standards**

Number and Operations
Students will use multiplication, division, factoring, and arithmetic to represent the cryptological processes they are examining.

Algebra
Students will use matrices and tables to develop mathematical constructs for the modeling processes of encryption and decryption.

Data Analysis and Probability
Students will organize and interpret numerical data gathered from the problem they are analyzing. Using this data, they also will develop and evaluate inferences, predictions, and arguments which are based on that data.

- **Process Standards**

Problem Solving
Students will use logic skills and the analysis of patterns and numerical data to solve enciphered messages.

Reasoning and Proof
Students will use inductive reasoning to devise mathematically sound approaches for decrypting messages.

Connections
Students will discover connections among the following branches of mathematics: algebra, numerical analysis, factoring, probability and statistics.

Representation
Students will learn to represent the processes of encryption and decryption using mathematical terms.

Links to Maryland High School Mathematics Core Learning Goals:

Functions and Algebra

- **1.1.1**
  Students will recognize and describe patterns to develop equations which model the processes of encryption and decryption.

- **1.1.3**
  Students will manipulate algebraic expressions in order to evaluate equations.

Geometry, Measurement, and Reasoning

- **2.2.3**
  Students will work with matrices and use inductive and deductive reasoning to develop methods of decryption.

Data Analysis and Probability

- **3.1.1**
  Students will use statistical methods to analyze frequency counts and indices of the cipher and resultant plaintext.

Grade/Level:

Grades 9-12

Duration/Length:

Two 45-minute periods

Prerequisite Knowledge:

Students should have working knowledge of the following skills:

- Basic arithmetic, multiplication, and division
- Algebra

Student Outcomes:

Students will:

- understand the differences between the two major families of encryption systems.
- learn the basic principles of transposition encipherment.
- learn to use data analysis to help solve problems.
- use pattern recognition to express processes in a mathematical fashion.

Materials/Resources/Printed Materials:

- Paper
- Pencil
- Calculator

Development/Procedures:

Lesson 1: Cryptanalysis of Matrix Transposition Cipher

The first step in any cryptanalysis is to get a frequency count of each of the cipher characters. An examination of this will show a wide discrepancy in distribution of the frequencies of the various letters. Such a distribution is referred to as a rough distribution. We have a formula which can give us a quantitative measure of the roughness of a distribution, as well as an indication of the number of cipher alphabets used in the encryption. This formula is called the index of coincidence (hereafter referred to as the i.c.) and it is calculated as follows:

$$\sum_{i=a}^{i=z} \frac{fi\,(fi - 1)}{N\,(N-1)}$$

This is a summation, where fi is the frequency of each letter and N is the total number of characters in the message. Index of coincidence values above 0.6 indicate a monoaphabetic substitution system. Calculating the i.c. of a transposed message will indicate the cipher is monoalphabetic. The monographic frequency count will show a pattern consistent with plaintext English. Obviously, the matrix can take the shape of any of the possible factors of the total number of cipher letters. Traditionally however, cryptographers have preferred to use square, or nearly square, shaped matrices for ease of operation. This fact will usually reduce the number of possibilities to one or two. The first R letters of cipher become the left-most column of the proposed matrix, where R equals the number of rows in the matrix. The second string of R letters becomes the next column; the third string of letters becomes the third column; and so on, until the matrix is filled.

When there is a choice of two or more possible matrix sizes, there must be a way to determine which, if any, is correct. Fortunately the English language has a mathematical property that can help. In normal writing, roughly 40% of the letters used are vowels. It therefore stands to reason that if plaintext is inscribed into a matrix, about 40% of the letters in any row will be vowels.

Consider the following cipher message:

ETTFF AETTT EHFES OTTDL THEEU IMTIS TRCEE EELSU ONHUN FEIOM
NHRAV MRENL BLPEN HITME SRPET MIFDT WCIYA EYTOO TCHES RCLNC
PESGR HFROD ISTUO LDEEN SOUEN OLEST STBSC FSIOT ROIJE NIMOR
DEEVO RDIOE EOTTE SIELO EEEUN TNROA NRAAP NNAUS UROOP NABIE
RHATT IAPED ROUDQ DOMAS SRESE ESNDR THEOE BEIYH NEATO UOYAI
IESWF AORTC SIRFT LRGOD IRBSS T (276)

```
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
==================================================
1 5 7 1 4 9 2 1 1 1 0 9 7 1 2 6 1 2 2 3 1 2 2 0 4 0
3     1 3   0 9           7 6     1 1 0 0
```

The total number of cipher characters is 276. Factoring this number gives us the possible matrix sizes of 12 rows X 23 columns or 23 rows X 12 columns.

	#VOWELS	#VOWELS/12
EEEEHLOOPPTW	6	.50
TUISEDTTNEHF	4	.33
TIORSERTNDEA	5	.42
FMMPREOEAROO	6	.50
FTNECNISUOER	5	.42
AIHTLSJISUBT	4	.33
ESRMNOEEUDEC	6	.50
TTAICUNLRQIS	4	.33
TRVFPEIOODYI	5	.42
TCMDENMEOOHR	4	.33
EERTSOOEPMNF	5	.42
HEEWGLRENAET	5	.42
FENCREDUASAL	5	.42
EELIHSENBSTR	4	.33
SLBYFTETIROG	3	.25
OSLARSVNEEUO	6	.50
TUPEOTORRSOD	5	.42
TOEYDBROHEYI	5	.42
DNNTISDAAEAR	5	.42
LHHOSCINTSIB	3	.25
TUIOTFORTNIS	5	.42
HNTTIUSEAIDET	6	.50
EFMCOIEAARST	5	.42

	#VOWELS	#VOWELS/23
ETTFFAETTTEHFESOTTDLTHE	7	.30
EUIMTISTRCEEEELSUONHUNF	11	.50
EIOMNHRAVMRENLBLPENHITM	7	.30
ESRPETMIFDTWCIYAEYTOOTC	8	.35
HESRCLNCPESGRHFRODISTUO	6	.26
LDEENSOUENOLESTSTBSCFSI	7	.35
OTROIJENIMORDEEVORDIOEE	12	.57

OTTESIELOEEEUNTNROANRAA	12	.57
PNNAUSUROOPNABIERHATTIA	11	.50
PEDROUDQDOMASSRESEESNDR	8	.35
THEOEBEIYHNEATOUOYAIIES	13	.61
WFAORTCSIRFTLRGODIRBSTT	5	.22

Examination of the cipher written into the two possible matrices and dividing by the number of columns shows that 23 rows X 12 columns more consistently approximates the expected value of 0.40 for plaintext. However, even though we are fairly confident we have correctly reconstructed the matrix shape, the message does not appear because we do not know the extraction key used by the cryptographer. That is to say, the columns are not in the correct order.

Once again we can use a property of the English language to assist us. That property is cohesion. Simply put, cohesion means that certain letters tend to be used together often. The best example of this is the pairing of QU.

In any English word that contains a Q, the next letter is always U, which is followed by another vowel. Referring to our monographic frequency table, we see that there is only one Q. It is located in the eight row, tenth column of the matrix. There is only one U in the same row, so these two columns must be adjacent . (Only the first ten rows will be written for brevity)

PL
ED
DE
RE
ON
US
DO
QU
DE
ON

The choice of the vowel to follow the U is empirical, so we will arbitrarily choose the I in the fourth column.

PLE
EDS
DER
REP
ONE
UST
DOM
QUI
DEF
OND

The UST in the sixth row could be MUST, TRUST, or JUST. The J in the seventh column is our only choice so we will move that column to the left side.

OPLE
TEDS
RDER
OREP
IONE
JUST
EDOM
NQUI
IDEF
MOND

The form forming on the third row looks like ORDER, so we move the third column to the left side.

EOPLE
ITEDS
ORDER
MOREP
NIONE
HJUST
REDOM
ANQUI
VIDEF
MMOND

The word on the first row must be PEOPLE. There are two P's in that row, but column nine is a better fit for the other rows in our recovery matrix.

PEOPLE
NITEDS
NORDER
AMOREP
UNIONE
SHJUST
UREDOM
RANQUI
OVIDEF
OMMOND

This process is called anagramming - the shifting of columns in the matrix to simultaneously recover plaintext words on each of the rows. This continues until all the columns have been replaced and the matrix is full.

Assessment:

Student progress can easily be measured by their deduction of the correct shape of the enciphering matrix and the recovery of the plaintext. This is a process that lends itself to either working in groups or individual effort by enthusiastic students.

Extension/Follow Up:

Students could work together in pairs to write, encrypt, and decrypt their own messages using the principles described in class.

Authors:

Math Education Partnership Program Office
National Security Agency
Fort Meade, MD